WHAT IS CLAIMED IS:

1. A content-based method for retrieving data files among a set of database files comprising:

providing positive and negative examples of data files; said positive example including at least one relevant feature;

providing at least one discriminating feature in at least one of said positive and negative examples allowing to differentiate between said positive and negative examples;

for each database file in said set of database files, computing a relevance score based on a similarity of said each database file to said positive example considering said at least one relevant feature;

creating a list of relevant files comprising the Nb_1 files having the highest similarity score among said set of database files; Nb_1 being a predetermined number;

for each relevant file in said list of relevant files, computing a discrimination score based on a similarity of said each relevant file to said positive example considering said at least one discriminating feature and on a dissimilarity of said each relevant file to said negative example considering said at least one discriminating feature; and

selecting the Nb_2 files having the highest discrimination score among said list of relevant files; Nb_2 being a predetermined number.

2. A content-based method for retrieving images among a set of database images comprising:

providing positive and negative example images; said positive example image including at least one relevant feature;

providing at least one discriminating feature in at least one of said positive and negative examples allowing to differentiate between said positive and negative example images;

for each database image in said set of database images, computing a relevance score based on a similarity of said each database image to said positive example image considering said at least one relevant feature;

creating a list of relevant images comprising the Nb_1 images having the highest relevance score among said set of database images; Nb_1 being a predetermined number;

for each relevant image in said list of relevant images, computing a discrimination score based on a similarity of said each relevant image to said positive example image considering said at least one discriminating feature and on a dissimilarity of said each relevant image to said negative example image considering said at least one discriminating feature; and

selecting the Nb_2 images having the highest discrimination score among said list of relevant images; Nb_2 being a predetermined number.

3. A method as recited in claim 2, wherein said at least one of said positive and negative examples being the weighted average of a plurality of images.

4. A method as recited in claim 2, wherein said at least one relevant feature includes a number *l* of relevant features.

5. A method as recited in claim 4, wherein said positive example image being the weighted average $\bar{x}_l$ of N_1 positive examples for each relevant feature *l*.

6. A method as recited in claim 5, wherein $\bar{\bar{x}}_i^1$ is defined by:

$$\bar{\bar{x}}_i^1 = \frac{\sum_{n=1}^{N_1} \pi_n^1 x_{ni}^1}{\sum_{n=1}^{N_1} \pi_n^1}$$

wherein π_n^1 is a relevance degree for the positive example *n*.

7. A method as recited in claim 6, wherein said at least one discriminating feature includes a number *I* of discriminating features; said negative example image being the weighted average $\bar{\bar{x}}_i^2$ of N_2 negative examples for each relevant feature *i*; $\bar{\bar{x}}_i^2$ being defined by:

$$\bar{\bar{x}}_i^2 = \frac{\sum_{n=1}^{N_2} \pi_n^2 x_{ni}^2}{\sum_{n=1}^{N_2} \pi_n^2}$$

wherein π_n^2 is a relevance degree for the negative example *n*.

8. A method as recited in claim 7, wherein $\tilde{\pi}^1 + \tilde{\pi}^2 = 1$ where: $\tilde{\pi}^1 = \sum_{n=1}^{N_1} \pi_n^1$ and $\tilde{\pi}^2 = \sum_{n=1}^{N_2} \pi_n^2$.

9. A method as recited in claim 8, wherein $\tilde{\pi}_1$=0.5 and $\tilde{\pi}_2$=0.5.

10. A method as recited in claim 2, wherein each of the set of database images and of the positive and negative example images is represented by a set of image features.

11. A method as recited in claim 3, wherein each of said set of image features being represented by a feature vector.

12. A method as recited in claim 11, wherein computing a relevance score includes computing the distance between said positive example image and said each database image; said highest relevance score corresponding to the lowest of said distance between said positive example image and said each database image.

13. A method as recited in claim 12, wherein said at least one relevant feature includes a number *I* of relevant features; said positive example image is the weighted average $\vec{\bar{x}}_i^1$ of N_1 positive examples for each relevant feature *i*; $\vec{\bar{x}}_i^1$ being defined by:

$$\vec{\bar{x}}_i^1 = \frac{\sum_{n=1}^{N_1} \pi_n^1 x_{ni}^1}{\sum_{n=1}^{N_1} \pi_n^1}$$

wherein π_n^1 is a relevance degree for the positive example *n;*

said distance between said positive example image and said each database image represented by feature vector $\vec{\bar{x}}_{ni}$ being defined by:

$$D(x_n) = \sum_{i=1}^{I} u_i (\vec{x}_{ni} - \vec{\bar{x}}_i^1)^T W_i (\vec{x}_{ni} - \vec{\bar{x}}_i^1)$$

wherein u_i is the global weight assigned to the i[th] relevant feature; and

W_i is a symmetric matrix that allows defining the generalized ellipsoid distance *D* and weighting components of each of said at least one relevant feature; and u_i and W_i minimizing the dispersion $J_{positive}$ of positive example images

$$J_{positive} = \sum_{i=1}^{I} u_i \sum_{n=1}^{N_1} \pi_n^1 (\vec{x}_{ni}^1 - \vec{\bar{x}}_i^1)^T W_i (\vec{x}_{ni}^1 - \vec{\bar{x}}_i^1)$$

14. A method as recited in claim 12, wherein computing a discrimination score includes computing the distance between said negative

example image and said each database image; said highest discrimination score corresponding to the lowest of said distance between said negative example image and said each database image.

15. A method as recited in claim 14, wherein said at least one relevant feature includes a number *I* of relevant features; said positive example image is the weighted average $\bar{\bar{x}}_i^1$ of N_1 positive examples for each relevant feature *i*; $\bar{\bar{x}}_i^1$ being defined by:

$$\bar{\bar{x}}_i^1 = \frac{\sum_{n=1}^{N_1} \pi_n^1 x_{ni}^1}{\sum_{n=1}^{N_1} \pi_n^1}$$

wherein π_n^1 is a relevance degree for the positive example *n;*

said negative example image is the weighted average $\bar{\bar{x}}_i^2$ of N_2 negative examples for each relevant feature *i*; $\bar{\bar{x}}_i^2$ being defined by:

$$\bar{\bar{x}}_i^2 = \frac{\sum_{n=1}^{N_2} \pi_n^2 x_{ni}^2}{\sum_{n=1}^{N_2} \pi_n^2}$$

wherein π_n^2 is a relevance degree for the negative example *n;*

said distance between said positive example image and said each database image represented by feature vector $\bar{\bar{x}}_{ni}$ minus said distance between said negative example image and said each database image represented by feature vector $\bar{\bar{x}}_{ni}$ being defined by:

$$D(x_n) = \sum_{i=1}^{I} u_i(\vec{x}_{ni} - \bar{\bar{x}}_i^1)^T W_i(\vec{x}_{ni} - \bar{\bar{x}}_i^1) - \sum_{i=1}^{I} u_i(\vec{x}_{ni} - \bar{\bar{x}}_i^2)^T W_i(\vec{x}_{ni} - \bar{\bar{x}}_i^2)$$

wherein u_i is the global weight assigned to the i^{th} relevant feature; and

W_i is a symmetric matrix that allows to define the generalized ellipsoid distance *D*; and u_i and W_i minimizing the internal dispersion of positive example images, minimizing the internal dispersion of the negative example

images, and maximizing the discrimination between the positive and the negative examples.

16. A method as recited in claim 15, wherein minimizing the internal dispersion of positive example images, minimizing the internal dispersion of the negative example images, and maximizing the discrimination between the positive and the negative examples is achieved by minimizing A/R where:

$$A = \sum_{i=1}^{I} u_i \sum_{k=1}^{2} \sum_{n=1}^{N_k} \pi_n^k (\vec{x}_{ni}^k - \vec{\bar{x}}_i^k)^T W_i (\vec{x}_{ni}^k - \vec{\bar{x}}_i^k)$$

$$R = \sum_{i=1}^{I} u_i \sum_{k=1}^{2} \tilde{\pi}^k (\vec{\bar{x}}_i^k - \vec{q}_i)^T W_i (\vec{\bar{x}}_i^k - \vec{q}_i)$$

where k = 1 for positive example and k = 2 for negative example, and where $\vec{q}_i$ is the weighted average of all positive and negative example images for the i^{th} feature and is defined by

$$\vec{q}_i = \frac{\sum_{k=1}^{2} \sum_{n=1}^{N_k} \pi_n^k \vec{x}_{ni}^k}{\sum_{k=1}^{2} \sum_{n=1}^{N_k} \pi_n^k}$$

17. A method as recited in claim 2, wherein said positive and negative example images are selected by a person among a list of sample images.

18. A content-based method for retrieving data files among a set of database files, the method comprising:

providing positive and negative example of data files; said positive example image including at least one relevant feature;

restricting the set of database files to a subset of files selected among said database files; each files in said subset of files being selected according to its similarity with said positive example based on said at least one relevant feature;

retrieving files in said subset of files according to their similarity with said positive example based on said at least one relevant feature and according to their dissimilarity with said negative example based on at least one discriminating feature between said positive and negative examples;

whereby, the files retrieved among said database files corresponding to files similar to said positive example and dissimilar to said negative example.

19. A content-based method for retrieving images among a set of database images, the method comprising:

providing positive and negative example images; said positive example image including at least one relevant feature;

restricting the set of database images to a subset of images selected among said database images; each images in said subset of images being selected according to its similarity with said positive example based on said at least one relevant feature;

retrieving images in said subset of images according to their similarity with said positive example based on said at least one relevant feature and according to their dissimilarity with said negative example based on at least one discriminating feature between said positive and negative examples;

whereby, the images retrieved among said database images corresponding to images similar to said positive example and dissimilar to said negative example.

20. A content-based system for retrieving images among a set of database images comprising:

means for providing positive and negative example images; said positive example image including at least one relevant feature;

means for providing at least one discriminating feature in at least one of said positive and negative examples allowing to differentiate between said positive and negative example images;

means for computing, for each database image in said set of database images, a relevance score based on a similarity of said each database image to said positive example image considering said at least one relevant feature;

means for creating a list of relevant images comprising the Nb_1 images having the highest similarity score among said set of database images; Nb_1 being a predetermined number;

means for computing, for each relevant image in said list of relevant images, a discrimination score based on a similarity of said each relevant image to said positive example image considering said at least one discriminating feature and on a dissimilarity of said each relevant image to said negative example image considering said at least one discriminating feature; and

means for selecting the Nb_2 images having the highest discrimination score among said list of relevant images; Nb_2 being a predetermined number.

21. A system as recited in claim 20, wherein said means for providing positive and negative example images includes a graphical user interface displaying sample images.

22. A system as recited in claim 20, wherein said graphical user interface includes means for specifying the degree of relevance of each said sample images.

23. A system as recited in claim 22, wherein said graphical user interface includes means for viewing the retrieved images.

24. An apparatus for retrieving images among a set of database images, the apparatus comprising:

an interface adapted to receive positive and negative example images; said positive example image including at least one relevant feature;

a restriction component operable to restrict the set of database images to a subset of images selected among said database images; said images in said subset of images being selected according to their similarity with said positive example based on said at least one relevant feature;

a retrieval component operable to retrieve images in said subset of images according to their similarity with said positive example based on said at least one relevant feature and according to their dissimilarity with said negative example based on at least one discriminating feature between said positive and negative examples;

whereby, the images retrieved among said database images correspond to images similar to said positive example and dissimilar to said negative example.

25. An apparatus according to claim 24, wherein the restriction component and the retrieval component are implemented within the same logic device.

26. A computer readable memory comprising content-based image retrieval logic for retrieving images among a set of database images, the content-based image retrieval logic comprising:

image reception logic operable to receive positive and negative example images; said positive example image including at least one relevant feature;

restriction logic operable to restrict the set of database images to a subset of images selected among said database images; said images in said subset of images being selected according to their similarity with said positive example based on said at least one relevant feature; and

retrieval logic operable to retrieve images in said subset of images according to their similarity with said positive example based on said at least one relevant feature and according to their dissimilarity with said negative example based on at least one discriminating feature between said positive and negative examples;

whereby, the images retrieved among said database images correspond to images similar to said positive example and dissimilar to said negative example.